

MODELING AVIAN ABUNDANCE FROM REPLICATED COUNTS USING BINOMIAL MIXTURE MODELS

MARC KÉRY,^{1,3} J. ANDREW ROYLE,² AND HANS SCHMID¹

¹Swiss Ornithological Institute, 6204 Sempach, Switzerland

²USGS Patuxent Wildlife Research Center, 12100 Beech Forest Rd., Laurel, Maryland 20708 USA

Abstract. Abundance estimation in ecology is usually accomplished by capture–recapture, removal, or distance sampling methods. These may be hard to implement at large spatial scales. In contrast, binomial mixture models enable abundance estimation without individual identification, based simply on temporally and spatially replicated counts. Here, we evaluate mixture models using data from the national breeding bird monitoring program in Switzerland, where some 250 1-km² quadrats are surveyed using the territory mapping method three times during each breeding season. We chose eight species with contrasting distribution (wide–narrow), abundance (high–low), and detectability (easy–difficult). Abundance was modeled as a random effect with a Poisson or negative binomial distribution, with mean affected by forest cover, elevation, and route length. Detectability was a logit-linear function of survey date, survey date-by-elevation, and sampling effort (time per transect unit). Resulting covariate effects and parameter estimates were consistent with expectations. Detectability per territory (for three surveys) ranged from 0.66 to 0.94 (mean 0.84) for easy species, and from 0.16 to 0.83 (mean 0.53) for difficult species, depended on survey effort for two easy and all four difficult species, and changed seasonally for three easy and three difficult species. Abundance was positively related to route length in three high-abundance and one low-abundance (one easy and three difficult) species, and increased with forest cover in five forest species, decreased for two nonforest species, and was unaffected for a generalist species. Abundance estimates under the most parsimonious mixture models were between 1.1 and 8.9 (median 1.8) times greater than estimates based on territory mapping; hence, three surveys were insufficient to detect all territories for each species. We conclude that binomial mixture models are an important new approach for estimating abundance corrected for detectability when only repeated-count data are available. Future developments envisioned include estimation of trend, occupancy, and total regional abundance.

Key words: abundance estimation; binomial mixture model; breeding bird surveys; count data; detectability; index of abundance; monitoring; random effect; replicated counts; Switzerland.

INTRODUCTION

The study of spatial and temporal variation in abundance is central to ecology (Krebs 2001). Yet, most species are so widespread or so inconspicuous that their abundance cannot be assessed without error, but instead must be estimated using methods that account for detectability. The estimation of abundance is of fundamental importance in both basic and applied ecology. In biological monitoring programs, in particular, abundance is an important state variable (Yoccoz et al. 2001, Pollock et al. 2002). Most monitoring programs use counts of organisms as proxies for true abundance. In so doing, they make the implicit assumption of a proportional relationship between count index and true abundance. Detectability p of the counted objects is assumed to be either perfect (i.e., $p = 1$), or at least its expectation is assumed to be constant across tem-

poral or spatial dimensions to be compared. The first assumption may be met only in exceptional cases, and the second is at least questionable and should be evaluated.

Most methods to obtain detectability-corrected abundance estimates involve some form of capture–recapture, removal, or distance sampling (Buckland et al. 2001, Williams et al. 2002). In capture–recapture sampling, information about the detectability of organisms is obtained from recapture (or resighting) information on individuals. In removal sampling, the frequency of removals (or first sightings) in successive sample periods contains detectability information, whereas in distance sampling, the distribution of detections in relation to a point or transect of observation yields information on detectability. Sometimes these approaches may be impractical, e.g., when individual identification is impossible, distance measurement unreliable, or when sample sizes for individual sampling units are small. This often has been used as justification for using “indices” based on counts rather than detectability-

Manuscript received 15 July 2004; revised 6 October 2004; accepted 22 October 2004; Final version received 9 December 2004. Corresponding Editor: T. R. Simons.

³ E-mail: marc.kery@vogelwarte.ch

corrected estimates of true abundance. Raw data resulting from simple count surveys are of little value due to the ambiguity induced by imperfect detection (e.g., Anderson 2001, Rosenstock et al. 2002).

Binomial mixture models are a class of models for estimating and modeling both abundance and detection probability from count data (Dodd and Dorazio 2004, Royle 2004a). They enable detectability-corrected abundance estimates in the absence of individual identification. The key requirement of these models is the temporal replication of counts at a number of sample locations. The modeling assumes that there are no changes in abundance over the survey period (i.e., demographic closure). Repeated counts within sample location i may then be viewed as independent realizations of a binomial random variable with parameters N_i (local abundance) and p_i (detection probability). It is further assumed that N_i comes from some common distribution specified by parameters to be estimated from the data. Both N_i and p_i may be modeled as functions of covariates to increase precision or to investigate covariate relationships.

Virtually all biological monitoring programs for birds collect simple count data, and their results will be biased to an unknown degree by heterogeneous and imperfect detection (Pollock et al. 2002). Therefore, mixture models seem to have considerable potential for unbiased abundance estimation. In this paper, we evaluate the utility of mixture models for a large-scale study, using data from the national monitoring program of common breeding bird species in Switzerland (Schmid et al. 2001, Kéry and Schmid 2004). We chose data from 2002 for eight species with contrasting distribution, density, and detectability, to confront the models with contrasting sample sizes, abundance, and detectability. We compare the models for two mixture distributions and several covariates on abundance and detectability.

METHODS

The site-by-survey matrix in biological surveys

The data arising from many large-scale programs designed to assess abundance can be summarized as a species-specific site-by-survey matrix of counts C . Columns j represent temporally repeated surveys. Rows i represent different spatial samples, i.e., sites, quadrats, or routes. Element c_{ij} is the number c of animals counted during survey j at site i . Typically, C will be sparse, i.e., many counts will be low or even zero. Many rows will contain all zeroes, representing quadrats where a species was not detected at all.

Numerical example

Study area.—We evaluated mixture models for abundance estimation using the national breeding bird monitoring program in Switzerland (Schmid et al. 2001). Switzerland is a small (41 285 km²), highly mountain-

ous country in Central Europe, with elevations ranging from 200 to 4600 m a.s.l. Median elevation of the surveyed quadrats in our study was 1180 m and ranged from 250 to 2750 m. At higher elevations, there are virtually no breeding bird species. Forests in most parts of the country are small and fragmented, and forest cover averages ~30%. Most nonforested areas below 600 m in elevation are either urban or used for small-scale, but intensive, agriculture. Human population density and the intensity of agricultural and forest land-use decrease with increasing elevation.

Field methods.—The Swiss monitoring program for common breeding bird species (“Monitoring Häufige Brutvögel,” or MHB) was launched in 1999 by the Swiss Ornithological Institute (Schmid et al. 2001, Kéry and Schmid 2004). More than 250 1-km² quadrats are distributed in a grid sample across Switzerland. During the breeding season (15 April–15 July), each quadrat is surveyed three times annually by an experienced observer along a quadrat-specific route using the territory mapping method (Bibby et al. 1992, Fewster et al. 2000). Only two surveys are conducted for ~50 quadrats above the alpine tree line at elevations greater than ~2000 m a.s.l. Routes aim to cover as large a proportion of a quadrat as possible and, once established, remain the same every year. During each survey, an observer marks every visual or acoustic contact with a potential breeding species on a large-scale map and notes additional information such as sex, behavior, territorial conflicts, location of pairs, or simultaneous observations of individuals from different territories. Date and time are also noted for each survey. For each quadrat and year, two or three repeated territory counts are available. Data on elevation and forest cover were taken from databases of the Swiss Federal Statistical Office. Previous analyses of data from the MHB have neglected the fact that the probability of detecting a territory will, in general, not equal 1, and may vary across space and time (Rosenstock et al. 2002, Diefenbach et al. 2003, Sauer et al. 2003). Hence, abundance of territories in a quadrat was estimated as the observed number of territories. Here, we present models that accommodate detection probability at the level of the individual territory, and that permit detectability-corrected estimates of abundance to be obtained for each species and quadrat.

Study species.—We chose survey data from 2002 for eight widely distributed species (Appendix A) with contrasting local distribution (narrow–wide), local abundance (low–high), and ease of detection at the level of an individual territory (easy–difficult): Mallard (*Anas platyrhynchos*), Hawfinch (*Coccothraustes coccothraustes*), Skylark (*Alauda arvensis*), Willow Tit (*Parus montanus*), Common Buzzard (*Buteo buteo*), Eurasian Jay (*Garrulus glandarius*), Blackbird (*Turdus merula*), and Coal Tit (*Parus ater*). None of these species are late-arriving migrants; all are breeding early in the season, and most territories are occupied from

TABLE 1. Model selection (AIC), goodness-of-fit (GoF) statistics, and estimates of abundance (\hat{N}_{total} and \hat{N}_{quadrat}) and detectability for three surveys combined (P^*) under binomial mixture models for eight bird species in the Swiss monitoring program for common breeding bird species (MHB).

			GoF	
Species and models	Distribution	AIC	SSE	<i>P</i>
Mallard				
Covariate models	Poi	506.8	61.53	0.006
	NB	472.7	60.15	0.104
Null models	Poi	631.9	86.89	<0.001
	NB	525.4	67.57	0.170
Hawfinch				
Covariate models	Poi	329.6	44.02	0.092
	NB	317.0	37.66	0.718
Null models	Poi	382.8	41.82	0.056
	NB†
Skylark				
Covariate models	Poi	1215.6	256.48	<0.001
	NB	862.9	197.29	0.418
Null models	Poi	1440.1	275.42	<0.001
	NB	900.9	201.04	0.368
Willow Tit				
Covariate models	Poi	1198.9	499.67	<0.001
	NB	1047.5	351.67	0.202
Null models	Poi	1784.6	468.25	<0.001
	NB	1161.9	358.23	0.266
Common Buzzard				
Covariate models	Poi	949.2	109.86	0.560
	NB	951.2	109.86	0.560
Null models	Poi	1053.9	120.70	0.178
	NB	1055.1	118.67	0.214
Eurasian Jay				
Covariate models	Poi	1536.1	423.02	0.058
	NB	1521.0	376.84	0.592
Null models	Poi	1777.7	437.92	<0.001
	NB	1669.6	365.75	0.426
Blackbird				
Covariate models	Poi	3600.5	3656.81	<0.001
	NB	3062.6	2616.78	0.716
Null models	Poi	5516.0	3835.25	<0.001
	NB	3376.7	3050.23	0.416
Coal Tit				
Covariate models	Poi	3559.2	3952.89	<0.001
	NB†
Null models	Poi	5182.8	4809.64	<0.001
	NB†

Notes: Results from the overall most parsimonious model are in boldface. We also give mean abundance estimated by the conventional territory mapping method (\hat{N}_{mapping}) and the ratio $\hat{N}_{\text{quadrat}}/\hat{N}_{\text{mapping}}$. Also shown are AIC, GoF statistics, and estimated total abundance $\hat{N}_{\text{total}} = \sum_{i=1}^R \hat{\lambda}_i$ under the most parsimonious mixture models with (covariate models) and without covariates (null models) and with Poisson (Poi) and negative binomial (NB) distribution. Further notation: SSE, sum-of-squared errors (bootstrap goodness-of-fit criterion); p , percentile of bootstrap distribution. See Appendix B for the covariates contained in the Poisson and negative binomial mixture models with covariates.

† Model unstable; no results given.

the start of the surveying period. We a priori judged the detectability of a territory based on our long-term field experience. In 2002, the eight species were detected at least once in between 32 (Hawfinch) and 197 1-km² quadrats (Blackbird). The mean number of territories detected per surveyed quadrat ($n = 238$ quadrats), i.e., the usual density estimate based on the territory mapping method, ranged from 0.22 (Hawfinch) to 14.42 (Blackbird; Table 1).

Mixture models

Up to now, it was believed that abundance and detection probability could not be separately estimated from simple point count data (e.g., Anderson 2001, 2003, Rosenstock et al. 2002). However, Royle (2004a) has developed a class of models that allows for estimation of both detection probability and abundance for the case wherein counts are replicated spatially and temporally within the context of a demographically

TABLE 1. Extended.

\hat{N}_{total} (CI)	Abundance			Detectability
	\hat{N}_{quadrat} (CI)	\hat{N}_{mapping}	Ratio	P_{\dagger} (SE)
91.0 (70–119)
103.7 (67–152)	0.43 (0.28–0.64)	0.41	1.05	0.88 (0.036)
82.7 (66–102)
112.5 (65–179)
77.4 (49–154)
432.8 (75–6103)	1.95 (0.32–25.64)	0.22	8.86	0.16 (0.266)
61.0 (40–95)
...
256.3 (223–297)
418.7 (265–824)	1.98 (1.11–3.46)	1.11	1.78	0.86 (0.040)
249.9 (219–280)
406.2 (250–668)
494.2 (414–616)
815.9 (441–1773)	3.43 (1.94–8.67)	1.89	1.81	0.60 (0.078)
347.6 (308–390)
900.0 (489–3321)
202.5 (168–240)	0.85 (0.71–1.01)	0.73	1.16	0.94 (0.014)
202.5 (168–240)
198.1 (164–230)
201.0 (167–240)
716.7 (570–987)
1019.4 (688–1800)	4.28 (2.89–7.56)	1.88	2.28	0.53 (0.065)
529.7 (452–639)
1079.1 (688–2159)
3500.4 (3173–3931)
7144.4 (5353–13 573)	30.60 (22.50–59.47)	14.42	2.12	0.66 (0.049)
2743.28 (2581–2922)
13 152.2 (10 649–16 439)
3189.0 (2815–3639)	13.40 (11.78–15.22)	11.21	1.20	0.83 (0.012)
...
2484 (2261–2721)
...

closed system, i.e., for surveys that yield a matrix of counts c_{ij} described previously.

Formally, let N_i be the local abundance for quadrat i . Demographic closure is manifest by the assumption that successive counts over the course of the study are binomial random variables with index N_i and detection probability p_{ij} . The model contains a large number of abundance “parameters” (one for each of a total of R sample locations); hence, the N_i are regarded as random effects with distribution $f(N_i; \theta)$. Estimation and inference are then focused on the parameter(s) θ . Although this view of local abundance as a random effect is motivated, in part, by the complexity of the model (i.e., the large number of abundance parameters), the view of local abundance as realizations of a random variable

can also be motivated by conventional metapopulation considerations (Royle 2004b). That is, we suppose that there exists a population of spatially indexed populations, each with size N_i . Then, our interest is in developing a characterization (i.e., model) of the metapopulation structure. This is, in effect, a goal that focuses on the parameter(s) θ .

A number of obvious choices of f are possible. Perhaps the most natural choice is to assume that N_i has a Poisson distribution with mean λ . The Poisson distribution is the customary description of a random spatial point pattern. In the case in which landscape covariates are available that explain variation in abundance, we consider the possibility that λ is site-specific, of the following form:

$$\log(\lambda_i) = \beta_0 + \sum_{k=1}^K x_{ik}\beta_k \quad (1)$$

where x_{ik} is the value of the k th covariate at site i . A natural generalization is to consider N_i to be negative binomial random variables (e.g., Lawless 1987). In this case, f is parameterized by an overdispersion parameter in addition to the mean. Again, models that include covariates may also be considered.

Detection probability may vary in response to covariates as well. To allow for this, we consider linear logistic models of the form

$$\text{logit}(p_{ij}) = \alpha_0 + \alpha_1 x_{ij}$$

for the case in which a single covariate, x_{ij} , exists (e.g., x_{ij} = survey duration).

Regardless of the model for N under consideration, estimation of abundance and detection probability parameters is based on the integrated likelihood, a standard approach for estimation and inference in classical random-effects models (e.g., Laird and Ware 1982). In the present context, first note that, conditional on N_i , the sampling distribution of the counts from site i is the product binomial:

$$g(\mathbf{c}_i | N_i, p) = \prod_j \text{Bin}(c_{ij} | N_i, p).$$

Without loss of generality, we have considered here the case where p is constant, but recognize that in most applications some covariates thought to influence p will be available. The marginal distribution of the counts, integrating over the distribution of N_i , is

$$g(\mathbf{c}_i | \theta, p) = \sum_{N_i = \max_j c_{ij}}^{\infty} \left[\prod_j \text{Bin}(c_{ij} | N_i, p) \right] f(N_i | \theta).$$

Regarded as a function of θ and p , $g(\mathbf{c}_i | \theta, p)$ is the contribution of the data from site i to the joint likelihood. Thus, the joint likelihood of the data from all R sites is the product

$$L(\mathbf{C} | \theta, p) = \prod_{i=1}^R g(\mathbf{c}_i | \theta, p).$$

Although this does not simplify algebraically in any meaningful way, it is simple enough to maximize using conventional numerical techniques in order to obtain the maximum likelihood estimates of the model parameters θ and p .

The maximum likelihood estimate of θ can be used to obtain an estimate of the expected number of territories in each quadrat based on the values of quadrat-specific covariates (Eq. 1). Royle (2004a) also describes an alternative estimator of N_i based on the posterior mean of N_i given the observed counts at quadrat i . Although this estimator based on the posterior mean is often similar to that based on Eq. 1, it is generally not under the negative binomial model.

Modeling strategy

We considered a Poisson or a negative binomial mixture distribution for abundance and several covariates for abundance and detectability. Due to the exceptional elevational gradient of Switzerland, all species have a distribution limit at some upper, and sometimes also at a lower, elevation. Furthermore, elevation is a surrogate for many habitat and climate covariates that influence distribution and abundance, e.g., human population density, agricultural land-use intensity, temperature, or precipitation. Therefore, we used elevation as a covariate for abundance N_i . We considered a linear and a quadratic elevation effect to take account of potential maximum abundance at medium elevations. Five species occur mostly in forests, two (Mallard and Skylark) in open land, and one (Common Buzzard) in both. We used percent forest cover as another habitat covariate. Route length is not standardized and varied slightly among quadrats, but was constant among surveys of each quadrat. To account for variation in effective sample area resulting from variable route length, we considered route length as a covariate for abundance.

Survey duration could vary among quadrats and among samples within each quadrat. We view survey duration as a potential influence on detectability: a higher proportion of available territories should be detected as sampling duration increases. However, because route length varies, a more appropriate covariate of sampling intensity is duration divided by route length, a measure of effort per sampled area. We refer to this covariate simply as effort. Because surveys were conducted over a period of up to three months, we expected considerable changes in singing behavior and other activities that facilitate territory detection (Royle and Nichols 2003, Selmi and Boulinier 2003). We therefore used survey date (1 = 1 April) as another covariate on detectability. We considered the possibility of a quadratic date effect to allow for nonlinear change in detectability due to seasonal variation in breeding behavior. Finally, we also fitted interaction terms between date and elevation as a covariate for detectability, because breeding takes place and surveys were conducted later at higher elevations. To enhance convergence of the numerical optimization algorithm, all covariates were standardized: route length was log-transformed and the remainder were transformed into standard normal deviates by first subtracting the arithmetic mean and then dividing by the standard deviation.

In total, 120 models were fit for each species and mixture type, containing all possible combinations of presence and absence for each covariate. We only considered models that conform to the usual marginality relations, e.g., no model containing an interaction without all constituent main effects was fit (Nelder 1994). We compared models using AIC and, within one mixture type, using ΔAIC , the distance in AIC units from

the most parsimonious model. As a rule of thumb, models with $\Delta AIC < 2$ fit a data set similarly well on the grounds of parsimony (Burnham and Anderson 1998). With the set of five most parsimonious models for each mixture type, we restricted discussion to those that are within 2 AIC units from the most parsimonious model. We called these the top models, and noted that different model sets could be considered, or that multimodel inference (model averaging; Buckland et al. 1997) would also be possible.

To define a detection-bias-adjusted index for abundance for the eight study species over all sampled quadrats in Switzerland, we considered the estimated abundance aggregated over the 238 sample units, that is, $\hat{N}_{total} = \sum_{i=1}^R \hat{N}_i$ where \hat{N}_i is the estimated posterior mean number of territories in quadrat i . Uncertainty about this summary of abundance was characterized using a parametric bootstrap procedure (Dixon 2002). For the negative binomial models, the resulting bootstrap distributions were highly skewed, so confidence intervals were used to summarize uncertainty about N_{total} . For comparison with this abundance estimate, we also give the conventional territory mapping abundance estimate, i.e., the total number of territories delimited when combining maps from all three surveys. In some studies, no meaningful covariates may be available. Then, a null model, with constant abundance and detectability may be useful. For comparison, we present results for a null model under both Poisson and negative binomial mixture distributions.

Goodness of fit

Adequacy of the best-fitting Poisson and negative binomial models was evaluated using a parametric bootstrapping procedure (Dixon 2002). For this procedure, parameters were fixed at the maximum likelihood estimates obtained for the model in question, and 500 replicate data sets were generated. For each replicate data set, parameters were estimated and a fit statistic (sum-of-squared errors in the present case) was computed. This collection of simulated values of the fit statistic forms the reference distribution to which the observed fit statistic is compared. Within this bootstrap goodness-of-fit framework, we also generated estimates of N_{total} for each replicate data set in order to characterize the sampling distribution of this statistic. This was used for purposes of evaluating its uncertainty, as described in the preceding section. Parameter estimation and goodness-of-fit assessment in this paper was achieved using the `nlm` function and our own code in the free software package R (Ihaka and Gentleman 1996). In a Supplement, we provide R code to conduct an example analysis of the Mallard data along with a tutorial.

RESULTS

Model selection and goodness of fit

For six of eight species, the negative binomial distribution provided a considerably more parsimonious

description of the data than did the Poisson, and was therefore preferred for inference (Table 1, Appendix B). Only the Common Buzzard data showed no overdispersion of abundance relative to a Poisson distribution. No numerical convergence could be achieved for negative binomial models in the Coal Tit; therefore, only results for Poisson mixtures are presented for that species. For five species, the data did not provide sufficient information for AIC to distinguish clearly between the five most parsimonious models; all were within 2 units from the AIC best model. Parameter estimates, model selection statistics, and mean abundance per 1-km² quadrat estimated under each of five most parsimonious models are shown in Appendix B.

Based on a parametric bootstrap, the negative binomial model usually fit adequately when it was indicated by AIC, and the corresponding Poisson model did not fit well. For the Common Buzzard, the Poisson mixture model did fit adequately (Table 1). For the Hawfinch and the Eurasian Jay, the most parsimonious Poisson model also (barely) fitted. The corresponding null models also provided a satisfactory fit each time when the respective covariate model also fit adequately.

Estimates of detectability and abundance

Estimates of total abundance for all 238 sampling quadrats combined (\hat{N}_{total}) were higher under the negative binomial than under the best Poisson covariate models (Table 1). Negative binomial confidence intervals were much wider and highly asymmetric, indicating considerable uncertainty about estimated abundance. Abundance estimates under the best null models had slightly wider confidence intervals than the corresponding models that contained covariates. Covariate information thus improved the precision of abundance estimates. Mean abundance per 1 km² ($\hat{N}_{quadrat}$) under the most parsimonious mixture model was between 1.05 and 8.86 (median 1.8) times greater than the conventional estimate ($\hat{N}_{mapping}$) based on the territory mapping method (Table 1). This reflects the fact that three surveys are not sufficient to detect all territories in each species. Except for the Hawfinch, where this discrepancy was greatest, abundance of most species was underestimated by the territory mapping method in the MHB by a factor of about 2 relative to mixture model estimates. Estimates of detectability combined for three surveys (P^*) under the most parsimonious model ranged from 0.66 to 0.94 (mean 0.84) for easy species, and from 0.16 to 0.83 (mean 0.53) for those species a priori judged to be difficult to detect (Table 1).

Covariate effects on abundance and detectability

Coefficients of the covariates present in the five most parsimonious negative binomial and Poisson models are shown in Appendix B. In the Mallard, abundance declined with increasing elevation as well as with increasing forest cover in all five top models. Surprisingly, in two of these models, there also was a negative

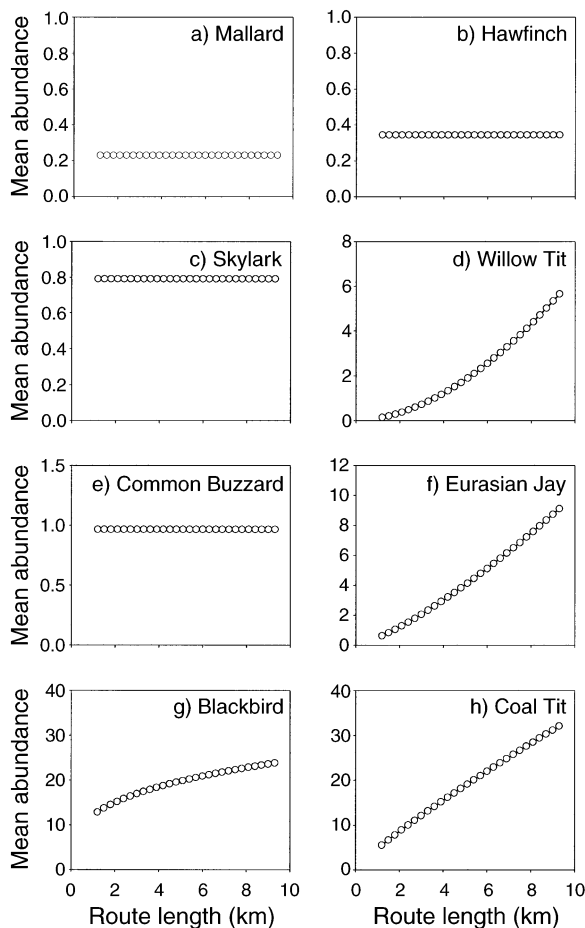


FIG. 1. Effect of route length on abundance estimates for eight study species in the Swiss breeding bird survey based on the most parsimonious model (Appendix B). The scale of the vertical axis differs among species.

route length effect, but the most parsimonious model did not contain this effect (Fig. 1a). Detectability declined over the season (Fig. 2a), and there were also negative date-by-elevation interactions. Detectability was not related to survey effort in the most parsimonious model (Fig. 3a).

For the Hawfinch, it was necessary to exclude from analysis one quadrat with exceptionally high territory counts (9–11) to achieve stable maximum likelihood estimates. All five top models contained a negative effect of elevation and a positive effect of forest cover. Two models, albeit not the most parsimonious one (Fig. 1b), contained a positive effect of route length on abundance. Detectability was not related to date (Fig. 2b), but was positively related to survey effort (Fig. 3b).

In the Skylark, all five top models contained negative effects on abundance of both elevation and forest cover. Again, two models, albeit not the most parsimonious one (Fig. 1c), contained a negative effect of route length. Three of the top five models (but not the most parsimonious, Fig. 2c) had a positive date effect and

one a date-by-elevation interaction. Detectability was positively affected by survey effort (Fig. 3c).

In the Willow Tit, all five top models had a positive effect of route length (Fig. 1d) and forest cover on abundance, and highest abundance at medium elevations. For detectability, there was a negative linear effect of date in all models and, in addition, two models contained either a quadratic effect of date (see Fig. 2d) or a date-by-elevation interaction. Three models had a positive effect of effort (Fig. 3d).

The Common Buzzard was the sole of eight studied species for which the Poisson models were more parsimonious than those with a negative binomial mixture for abundance. All five top models had negative linear and quadratic elevation effects on abundance. No effect on abundance was found for route length (Fig. 1e) or forest cover. Detectability declined during the season (Fig. 2e) and there was also a date-by-elevation interaction. One among the five top models (not the most

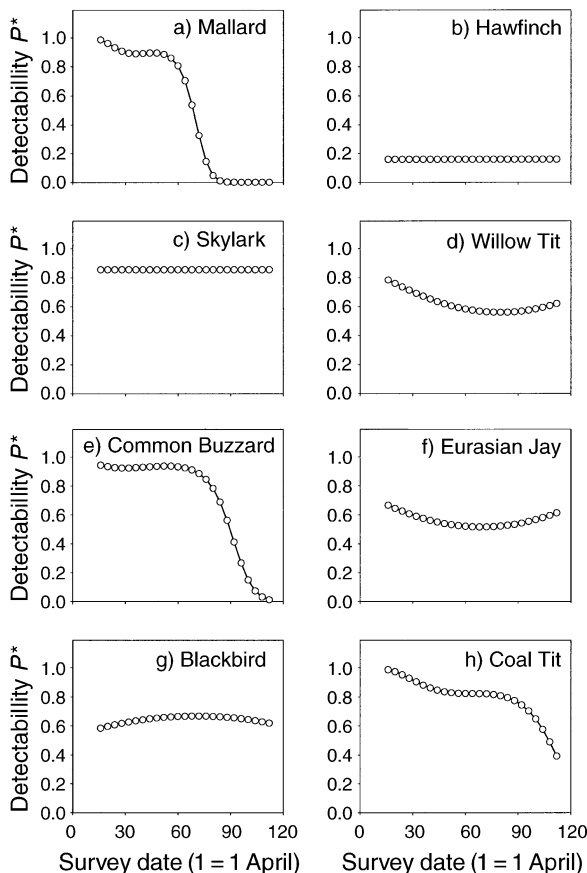


FIG. 2. Effect of date on per-year territory detectability estimates for eight study species in the Swiss breeding bird survey. Per-year detectability (the combined detectability over three surveys) was computed as $P^* = 1 - (1 - p)^3$, where p is the per-survey detectability from the most parsimonious model. We use P^* to illustrate how date effects would combine if all three surveys were conducted at the same date. Date 1 equals 1 April.

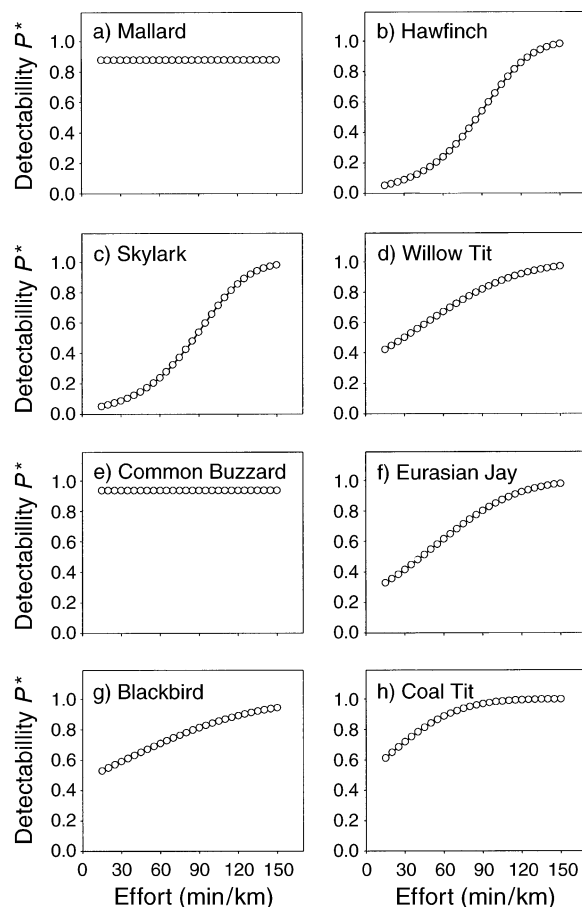


FIG. 3. Effect of survey effort on per-year territory detectability estimates for eight study species in the Swiss breeding bird survey computed as in Fig. 2. Effort is expressed as survey duration (in minutes) per route length (in kilometers).

parsimonious) also contained a positive effect of effort on detectability (Fig. 3e).

In the Eurasian Jay, only four models were within 2 AIC units of the most parsimonious model. They all contained positive effects on abundance of route length (Fig. 1f) and forest cover, and both linear and quadratic negative effects of elevation. Detectability declined over the season, with some modifications by elevation (Fig. 2f), and was positively related to survey effort (Fig. 3f).

In the Blackbird, the two models within 2 AIC units of the most parsimonious model only differed by the presence or absence of a date-by-elevation interaction. There were positive effects of route length (Fig. 1g) and forest cover on abundance, and negative effects of elevation. Detectability was highest in the middle of the season (Fig. 2g) and was positively related to survey effort (Fig. 3g).

For the Coal Tit, we were unable to obtain maximum likelihood estimates for negative binomial model parameters; we present results from a Poisson mixture

only. The single most parsimonious model contained effects of all covariates, and ranked 13.38 AIC units ahead of the second best model. There were positive effects on abundance of route length (Fig. 1h) and forest cover. Abundance was highest at medium elevations. Detectability declined over the season (Fig. 2h). In addition, there were also date-by-elevation interactions. Detectability was positively related to survey effort (Fig. 3h).

DISCUSSION

We evaluated mixture models (Royle 2004a) for obtaining detectability-corrected estimates of abundance at a large scale, and of covariate effects on abundance of territories and detectability, for eight bird species from the Swiss national breeding bird monitoring program (MHB). We chose species with contrasting distribution, abundance, and intrinsic detectability. We identified covariate effects and obtained estimates of detectability and abundance that were largely consistent with our expectations.

Biological observations

Mean territory detectability for three surveys combined (P^*) was estimated at 0.84 for easy species and 0.53 for difficult species. This means that in about two in 10 territories for easy species, and one in two territories for difficult species, birds were never seen or heard during three territory mapping surveys. Factors likely to affect detectability are bird size and habitat. Less detectable species were smaller, on average, and lived in denser habitats. In addition, they probably differ in behavior and perhaps territory size. The Hawfinch is a classic example of an elusive species that spends much time in the canopy, lacks striking vocalizations, and has a large home range. Consequently, its detectability was particularly low.

As expected, we found declining detectability over the season. This may reflect the fact that territorial activity is greatest at the beginning of the nesting season and then gradually declines, as birds become more involved in incubation and feeding offspring, rather than singing and territorial defense (Selmi and Bouli nier 2003).

Performance evaluation of mixture models under field conditions

Mixture models are known to yield unbiased estimates of abundance and detectability with simulated data chosen to conform to some defined distribution, and when populations are truly closed (Royle 2004a). In a field study, truth is unknown and there is no real benchmark with which to compare the performance of mixture models with that of conventional territory mapping. Abundance estimates under mixture models were between 1.1 and 8.9 (median 1.8) times greater than conventional territory mapping estimates. This

discrepancy may reflect bias of conventional methods and/or bias of mixture models.

We base our performance evaluation of mixture models on goodness-of-fit statistics, estimates of covariate coefficients, and of the magnitude of detectability and abundance estimates. Each model that was selected by AIC had an adequate goodness of fit. However, confidence intervals for the negative binomial models were wide. Where both the Poisson and the negative binomial model had adequate goodness of fit, one might estimate abundance under the Poisson model, even when it was not preferred by AIC.

The presence of effects in the top models, and their directions, concurred with expectations for each of the eight species. Significant elevation effects on abundance reflected the decline of abundance toward higher elevations shown for the same species by Schmid et al. (1998). Positive effects of route length on abundance were discernible for five species. As expected, the effect of forest cover was positive for five forest species, negative for Mallard and Skylark and neutral for the Common Buzzard. Detectability was higher for easy than for difficult species, and a positive effect of survey effort was present among the top five models for every species. Detectability varied over the season for six species in a direction consistent with a decline of territorial activity over the breeding season.

Our estimates of detectability and abundance estimates are hard to compare with other studies, because most published density estimates are of unknown quality and refer to small and often high-density areas (see, e.g., Glutz von Blotzheim and Bauer 1997). As expected, abundance estimates from mixture models were always higher than those obtained by territory mapping. They do not appear to be exaggerated except at first sight for the Hawfinch, with an estimated detectability of 0.16 ± 0.27 (mean \pm SE); almost nine times more territories were estimated under the negative binomial mixture than were identified using conventional territory mapping. However, even with 10–12 very thorough surveys, including nest searches by specialist observers, 22% of the pairs present may be missed (Glutz von Blotzheim and Bauer 1997:1206–1207). Using a simple binomial argument, detectability for three specialist surveys may be estimated at 0.34. This is well within 1 SE of our estimate, which is moreover based on data from a generalist survey aimed at counting ~ 100 species. For Willow and Coal Tit, a previous study estimated an even lower detectability for three surveys than we did: 30% for Willow Tit and 67% for Coal Tit (data from Blana [1978]). Three territory mapping surveys failed to detect about half of the territories for three marshland species (data from Bell et al. [1973]) and 35% for 18 woodland species (data from Blana [1978]). Underestimation of abundance by a factor 2 then appears to be reasonable for the territory mapping method with three surveys when detectability is ignored.

Choice of mixture distribution

There were substantial differences in abundance estimates, depending on the distribution assumed for the abundance random effect (see the last column in Appendix B); hence, the proper specification of that distribution is important for obtaining valid abundance estimates. Negative binomial mixtures were preferred for all except one species using information-theoretic methods (AIC). Common Buzzard abundance appeared to be consistent with a Poisson model at the scale of the 1-km² sampling units of the Swiss breeding bird monitoring. This species has large territories and most quadrats contain one territory at most. In contrast, all other species appear to exhibit excess variation in abundance relative to the Poisson. Bootstrap goodness-of-fit tests were consistent with model selection results based on AIC. Thus, both model selection and goodness-of-fit results indicate that abundance may be validly estimated under the models selected.

Nevertheless, the negative binomial distribution may not always be an ideal choice for representing the overdispersion in abundance relative to the Poisson. We feel that the mean/variance relationship implied by the negative binomial model may be extreme, and this has an especially deleterious effect on estimation for highly abundant species and those for which considerable variation is indicated. In such cases, the model fits the high mean and accommodates low counts by inflating the variance, yielding a strongly right-skewed abundance distribution that places considerable mass at excessive, and perhaps unrealistic, values of abundance.

Other random-effects distributions may be adopted for abundance that may prove adequate for describing overdispersion and may provide more stable maximum likelihood estimates than the negative binomial. We considered finite mixtures of Poisson distributions with some success. We also have briefly investigated the uniform-integrated Poisson (Bhattacharya and Holla 1965) and Generalized Hermite distributions (Puig 2003), but have not developed adequate experience with these at this time (J. A. Royle, *unpublished data*). More work on comparing different kinds of mixtures will be beneficial.

Covariate modeling

Variation in abundance may be accommodated implicitly as overdispersion, such as by the negative binomial distribution, or explicitly by modeling covariates into the mean. Although binomial mixture models worked well both for species with fairly restricted range and for those that are widespread (representing small and large sample sizes), we encountered some numerical problems for two extreme data sets (Hawfinch and Coal Tit). With more informative explanatory variables, it is likely that improvements could be achieved. Moreover, precision for Poisson models was much greater than for the negative binomial models (see nar-

lower CI values of the Poisson best model in Table 1). Obtaining an adequately fitting Poisson model by incorporating informative covariates would be very beneficial.

We used only a few environmental covariates that were likely to explain variation in the abundance of each species. Because of a representative quadrat sample in MHB, covariate modeling is not necessary for unbiasedness, but it still improves precision of the estimates (Table 1). In analyses of data that do not come from a random spatial sample (for instance, the North American Breeding Bird Survey; Sauer et al. 2003), covariate modeling is an attempt to correct for sample selection bias. In other situations, covariates on abundance may be chosen more specifically for just one or a few species. More precise abundance estimates may be possible for species about which more is known about the factors governing their occurrence and abundance.

The choice of which and how many covariates to introduce in an analysis depends on the goal of the analysis. For the sake of simplicity and comparability, complex covariate modeling may not be useful in large-scale multispecies studies. In this case, even null models may be useful (Table 1). In contrast, in single-species studies, reasonable gains in precision may be achievable at little cost when useful covariate information is used. In addition, extensive covariate modeling is more likely in studies that test for habitat relationships with abundance, for example.

Abundance estimates under different covariate models were fairly similar in our study (Appendix B). Where this is not the case, inference may be based on model averaging (Buckland et al. 1997) to properly account for model uncertainty.

Design issues

Detectability was <1 for all eight study species in a monitoring program based on the territory mapping method. In addition, detectability was also heterogeneous over time (date) and space (date-by-elevation interaction). Two key assumptions of conventional monitoring programs were clearly violated: that detectability is perfect, or at least that its average is constant over time and space. This has also been found by previous studies (e.g., Boulinier et al. 1998, Diefenbach et al. 2003, Selmi and Boulinier 2003). Barring double-counts, it means that abundance of all species will be underestimated systematically unless detectability is accounted for. Furthermore, estimates of relative abundance over time (trends) or space (when comparing areas) may be biased in studies that use raw counts instead of detectability-corrected estimates. This may not apply to local, intensive studies using territory mapping, when at just a few small sites, many more surveys (e.g., 8–15, Bell et al. 1973, Svensson 1978) are usually conducted in a breeding season. In exceptional cases, it may then be possible locally to truly

census a species, i.e., to detect every territory. For large-scale programs, however, this is unlikely ever to be the case, because limited resources must be distributed among many sites.

There were clear relationships between route length and abundance, and between survey effort and detectability. These are nuisance variables that may bias estimates when not accounted for. They can be eliminated in a statistical way by covariate modeling, but it may be preferable to reduce their effects by partial standardization at the design stage of a monitoring program. However, large-scale monitoring programs may depend on the work from hundreds of volunteers, and participation may be impaired by the imposition of stricter field protocols. There is a tension between rigor of design and number of available volunteers. To achieve greater sampling coverage in space and time, some lack of standardization may have to be accepted (but see Sauer et al. 2003). Further factors limiting the opportunity for standardization may be difficult on mountainous terrain such as in Switzerland, and with different optimal route lengths for different species. Therefore, covariate modeling may be an efficient remedy for nuisance variables.

For Mallard and Skylark, there were surprising negative effects of route length on abundance in some of the five top models (albeit not in the most parsimonious one). This counterintuitive result may stem from the fact that, in quadrats with good Mallard or Skylark habitat, shorter routes were chosen by field workers. More wetlands and open water may mean more Mallards, but perhaps also may reduce the accessibility of a quadrat. Skylarks attain highest densities in open farmland quadrats that are easy to survey and where observers select shorter than average routes. Our covariates may not be independent from hidden habitat variables. As a hypothetical example, different detectability of a species in different habitat might show up as an effect of survey effort on detectability in our analysis, when, in fact, walking speeds are affected by the same habitat types. This is a well-known complication for the causal interpretation of correlations.

For individual species, it would be possible to compute the optimal timing of a survey, i.e., when detectability is greatest. In most of the species that we examined, there was a monotonous decline in detectability over the season. However, there might be more well-defined optima in detectability for migratory species with a peak later in the season. Such information may be interesting for planning more focused surveys. In a multispecies study such as the MHB, detection of ~100 species would need to be optimized, there are early and late nesters, and there may be less opportunity to tune the design to a particular species.

In our study, the closure assumption was probably not violated because we selected species whose breeding populations are present over the entire survey period. In general, however, the closure assumption may

be critical for successful application of mixture models for counts. For migratory species, for instance, it might be fulfilled by deleting the data for those surveys that were conducted when a population was still arriving.

In this paper, we use count data to derive detectability-corrected abundance estimates. Arguably, counts provide less information on abundance and detectability than would data from a comparable design that also collects information on identity of individual territories or even birds. Hence, identity data ought to be collected whenever possible. However, especially at high densities, the identity of a territory or a bird might be difficult to ascertain without marking one of the territory owners, and this would be prohibitive at larger scales. Hence, we envision that, especially for high-density situations, mixture models based on counts are very competitive compared to other rigorous methods of abundance estimation. Examples might be counts of waterfowl or of common mammals such as hares.

Future developments

A key interest in monitoring lies in detecting temporal change in abundance (Dixon et al. 1998, Pollock et al. 2002). The current mixture model can be used with year as a factor (e.g., Dodd and Dorazio 2004) and the equality of abundance in certain years can be tested. Alternatively, annual change in abundance could be directly incorporated into the model as a parameter to be estimated, for instance as in $N_{it} \sim \text{Poisson}(\lambda_{it})$, with $\log(\lambda_{it}) = \alpha + \beta \times t$, where N_{it} is abundance for survey i in year t , λ_{it} is the estimated mean abundance for survey i in year t , t is a covariate representing time, and α and β are the intercept and slope parameter, respectively, of a linear trend in abundance. A comparison between two models with $\beta = 0$ and $\beta \neq 0$ represents a test for a significant trend. Accommodating trend as a parameter allows one to model it directly as a function of covariates, e.g., to test for habitat-specific or regional differences in trends. In addition, we are investigating open metapopulation models for species with imperfect detectability, where variation in abundance is partitioned into metapopulation dynamic components of extinction and colonization.

Occupancy is a special case of abundance; a species occurs at each site i where $N_i > 0$ (He and Gaston 2003, Royle and Nichols 2003). A characterization of the distribution of abundance is sufficient to also estimate occupancy. We show elsewhere (Royle et al. 2005) how mixture models for abundance allow estimates of the probability of occupancy. These estimates are free of the distorting effects of detectability, unlike conventional generalized linear model approaches (e.g., logistic regression), and they do not require restrictive assumptions about equality of abundance across sampling units, unlike the models of MacKenzie et al. (2002). Mapping a covariate model function for abundance or occupancy onto a wider area enables maps of potential abundance or range to be produced

(Royle et al. 2005). Integrating the volume under such an abundance distribution yields an estimate of the total population size over a larger area.

Conclusion

Mixture models have performed well for estimation of avian abundance on a large domain in this study. They hold promise for abundance estimation at large spatial scales because they require data that are easier to collect than those for previous methods that accommodate detectability. Key to their application is replication of counts in both time and space, and an adequate specification of the random-effects distribution of abundance. How counts are obtained is unimportant; counts of nests, wintering waterbirds, singing birds at point locations (point counts), or counts along transects, as well as the number of bird or plant species at repeated locations in species richness applications might all be used. They may be particularly useful when the counted objects occur at high densities. Future directions of research will include the direct incorporation of time trends into the modeling, the combined modeling of abundance and distribution, goodness-of-fit tests, and additional useful random-effects distributions for abundance.

ACKNOWLEDGMENTS

We thank Lukas Jenni, Romain Julliard, Verena Keller, Benedikt Schmidt, Niklaus Zbinden, Ted Simons, and two anonymous referees for valuable comments on earlier versions of this paper. We also warmly thank all volunteers who annually devote their valuable time to conduct the surveys in Switzerland. Roland Gautier extracted the data from data sheets. M. Kéry was partly funded by grant 81ZH-64044 from the Swiss National Science Foundation.

LITERATURE CITED

- Anderson, D. R. 2001. The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin* **29**:1294–1297.
- Anderson, D. R. 2003. Response to Engeman: index values rarely constitute reliable information. *Wildlife Society Bulletin* **31**:288–291.
- Bell, B. D., C. K. Catchpole, K. J. Corbett, and R. J. Hornby. 1973. The relationship between census results and breeding populations of some marshland passerines. *Bird Study* **20**:127–140.
- Bhattacharya, S. K., and M. S. Holla. 1965. On a discrete distribution with special reference to the theory of accident proneness. *Journal of the American Statistical Association* **60**:1060–1066.
- Bibby, C. J., N. D. Burgess, and D. Hill. 1992. *Bird census techniques*. Academic Press, London, UK.
- Blana, H. 1978. Die Bedeutung der Landschaftsstruktur für die Vogelwelt. *Beiträge zur Avifauna des Rheinlandes* **12**:1–225.
- Boulinier, T., J. D. Nichols, J. R. Sauer, J. E. Hines, and K. H. Pollock. 1998. Estimating species richness: the importance of heterogeneity in species detectability. *Ecology* **79**:1018–1028.
- Buckland, S. T., D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas. 2001. *Introduction to distance sampling. Estimating abundance of biological populations*. Oxford University Press, Oxford, UK.

- Buckland, S. T., K. P. Burnham, and N. H. Augustin. 1997. Model selection: an integral part of inference. *Biometrics* **53**:613–618.
- Burnham, K. P., and D. R. Anderson. 1998. Model selection and inference. Springer-Verlag, New York, New York, USA.
- Diefenbach, D. R., D. W. Brauning, and J. A. Mattice. 2003. Variability in grassland bird counts related to observer differences and species detection rates. *Auk* **120**:1168–1179.
- Dixon, P. M. 2002. Bootstrap resampling. *Encyclopedia of Environmetrics* **1**:212–220.
- Dixon, P. M., A. R. Olsen, and B. M. Kahn. 1998. Measuring trends in ecological resources. *Ecological Applications* **8**: 225–227.
- Dodd, C. K., Jr., and R. M. Dorazio. 2004. Using counts to simultaneously estimate abundance and detection probabilities in a salamander community. *Herpetologica* **60**:468–478.
- Fewster, R. M., S. T. Buckland, G. M. Siriwardena, S. R. Baillie, and J. D. Wilson. 2000. Analysis of population trends for farmland birds using generalized additive models. *Ecology* **81**:1970–1984.
- Glutz von Blotzheim, U. N., and K. M. Bauer. 1997. Handbuch der Vögel Mitteleuropas. Volume 14. Passeriformes (5. Teil). Aula-Verlag, Wiesbaden, Germany.
- He, F., and K. J. Gaston. 2003. Occupancy, spatial variance, and the abundance of species. *American Naturalist* **162**: 366–375.
- Ihaka, R., and R. Gentleman. 1996. R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5**:299–314.
- Kéry, M., and H. Schmid. 2004. Monitoring programs need to take into account imperfect species detectability. *Basic and Applied Ecology* **5**:65–73.
- Krebs, C. J. 2001. Ecology: the experimental analysis of distribution and abundance. Addison Wesley, Boston, Massachusetts, USA.
- Laird, N. M., and J. H. Ware. 1982. Random-effects models for longitudinal data. *Biometrics* **38**:963–974.
- Lawless, J. F. 1987. Negative binomial and mixed Poisson regression. *Canadian Journal of Statistics* **15**:209–225.
- MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. Estimating site occupancy when detection probabilities are less than one. *Ecology* **83**:2248–2255.
- Nelder, J. A. 1994. The statistics of linear models: back to basics. *Statistics and Computing* **4**:221–234.
- Pollock, K. H., J. D. Nichols, T. R. Simons, G. L. Farnsworth, L. L. Bailey, and J. R. Sauer. 2002. Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics* **13**:105–119.
- Puig, P. 2003. Characterizing additively closed discrete models by a property of their maximum likelihood estimators, with an application to generalized Hermite distributions. *Journal of the American Statistical Association* **98**:687–692.
- Rosenstock, S. S., D. R. Anderson, K. M. Giesen, T. Leukering, and M. F. Carter. 2002. Landbird counting techniques: current practices and an alternative. *Auk* **119**:46–53.
- Royle, J. A. 2004a. N-mixture models for estimating population size from spatially replicated counts. *Biometrics* **60**: 108–115.
- Royle, J. A. 2004b. Generalized estimators of avian abundance from spatially replicated count data. *Animal Biodiversity and Conservation* **27**:375–386.
- Royle, J. A. and J. D. Nichols. 2003. Estimating abundance from repeated presence-absence data or point counts. *Ecology* **84**:777–790.
- Royle, J. A., J. D. Nichols, and M. Kéry. 2005. Modeling occurrence and abundance of species with imperfect detection. *Oikos*, in press.
- Sauer, J. R., W. A. Link, and J. D. Nichols. 2003. Estimation of change in populations and communities from monitoring survey data. Pages 227–253 in D. E. Busch and J. C. Trexler, editors. *Monitoring ecosystems: interdisciplinary approaches for evaluating ecoregional initiatives*. Island Press, Washington, D.C., USA.
- Schmid, H., M. Burkhardt, V. Keller, P. Knaus, B. Volet, and N. Zbinden. 2001. Die Entwicklung der Vogelwelt in der Schweiz. Avifauna Report Sempach 1, Annex. Schweizerische Vogelwarte. Sempach, Switzerland.
- Schmid, H., R. Luder, B. Naef-Daenzer, R. Graf, and N. Zbinden. 1998. Schweizer Brutvogelatlas. Verbreitung der Brutvögel in der Schweiz und im Fürstentum Liechtenstein 1993–1996. Schweizerische Vogelwarte. Sempach, Switzerland.
- Selmi, S., and T. Boulinier. 2003. Does time of season influence bird species number determined from point-count data? A capture-recapture approach. *Journal of Field Ornithology* **74**:349–356.
- Svensson, S. E. 1978. Census efficiency and number of visits to a study plot when estimating bird densities by the territory mapping method. *Journal of Applied Ecology* **16**: 61–68.
- Williams, B. K., J. D. Nichols, and M. J. Conroy. 2002. Analysis and management of animal populations. Academic Press, San Diego, California, USA.
- Yoccoz, N. G., J. D. Nichols, and T. Boulinier. 2001. Monitoring of biological diversity in space and time. *Trends in Ecology and Evolution* **16**:446–453.

APPENDIX A

A table showing occupancy and preferred habitat of the eight study species is available in ESA's Electronic Data Archive: *Ecological Archives* A015-040-A1.

APPENDIX B

A table showing parameter estimates, number of parameters, model selection statistics (AIC, Δ AIC), and estimated mean abundance per surveyed 1-km² quadrat under the five most parsimonious models with a negative binomial or Poisson mixture distribution of abundance is available in ESA's Electronic Data Archive: *Ecological Archives* A015-040-A2.

SUPPLEMENT

Two files for fitting binomial mixture models as described originally by Royle (2004a) are available in ESA's Electronic Data Archive: *Ecological Archives* A015-040-S1.